Comparative Analysis of Statistical, Machine Learning, and Transformer-based Deep Learning Models for Climate-driven Crop Yield Prediction in Maharashtra

OPEN ACCESS

Volume: 13

Special Issue: 2

Month: January

Year: 2026

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Citation:
Ghawali, Ketaki, et al. "Comparative Analysis of Statistical, Machine Learning, and Transformer-Based Deep Learning Models for Climate-Driven Crop Yield Prediction in Maharashtra." *Shanlax International Journal of Arts, Science and Humanities*, vol. 13, no. 2, 2026, pp. 72–87.

DOI:
https://doi.org/10.34293/sijash.v13iS2-i1-Jan.10466

Ketaki Ghawali
Research Scholar, Assistant Professor
Mangalayatan University

Abhishek Garg
Associate Professor, Mangalayatan University

Mandar Sohani
Professor, VIT

Abstract
Maharashtra's agricultural productivity is increasingly threatened by climate change, manifested through shifting temperature regimes, altered rainfall patterns, and a growing frequency of extreme weather events. Economically significant crops such as sugarcane, cotton, and soybeans are particularly sensitive to these climatic perturbations, necessitating reliable data-driven approaches for climate-resilient planning and yield forecasting. This study presents a comparative evaluation of climate-driven crop yield prediction models at the district level in Maharashtra. A multi-year dataset integrating historical climatic variables—including temperature, rainfall, humidity, and solar radiation—was employed. Five predictive approaches were examined: ARIMA as a statistical baseline, Random Forest and XGBoost as machine learning models, Long Short-Term Memory (LSTM) networks for sequential learning, and the Temporal Fusion Transformer (TFT) to capture long-term dependencies through attention mechanisms. All models were trained and tested under identical experimental conditions. Performance was evaluated using ROC-AUC alongside regression-based metrics. Results indicate that machine learning and deep learning models consistently outperform the statistical baseline. Among them, TFT achieved the highest accuracy across all crops, followed by LSTM and XGBoost. These findings underscore the value of attention-based architectures for modeling climate–crop interactions and offer practical guidance for AI-driven agricultural decision-making in climate-vulnerable regions.
Keywords: Climate Change, Crop Yield Prediction, Machine Learning, Deep Learning, Temporal Fusion Transformer, Precision Agriculture

Introduction

Agriculture remains a central pillar of economic activity and rural livelihoods in many regions worldwide, particularly in developing economies such as India. Accurate prediction of crop yields has become increasingly critical due to the accelerating impacts of climate change on agricultural productivity. Climatic variability—including temperature extremes, altered precipitation regimes, and increased frequency of extreme weather events—presents significant challenges in forecasting crop outcomes (Betew et al., 2025; Lobell et al., 2011; Ray et al., 2015; Challinor et al., 2014; Schlenker & Roberts, 2009). In Maharashtra, one of India's most agriculturally dynamic states, these impacts are especially pronounced for major commercial crops such as sugarcane, cotton, and soybeans, which collectively support millions of farmers and contribute substantially to the regional agrarian economy. Precise yield forecasting at the district level would enable more robust planning, resource allocation, and mitigation strategies against climate-driven uncertainties (Lobell & Burke, 2010; Asseng et al., 2013; Rosenzweig et al., 2014).

Advanced analytical methods, spanning from classical statistical techniques to modern machine learning (ML) and deep learning (DL) frameworks, are rapidly gaining traction for crop yield prediction (van Klompenburg et al., 2020; Shawon et al., 2024; Wang et al., 2024). While traditional time series models such as AutoRegressive Integrated Moving Average (ARIMA) have long served as baselines for trend analysis, they often struggle to capture complex nonlinear dependencies inherent in climate–crop systems (Betew et al., 2025; Box et al., 2015). In contrast, machine learning models like Random Forest and XGBoost have demonstrated superior performance by effectively handling high-dimensional data and nonlinear relationships, integrating environmental covariates such as rainfall, temperature, and vegetation indices into predictive frameworks (Nikhil et al., 2024; Breiman, 2001; Chen & Guestrin, 2016). Studies have shown that ensemble learning approaches can achieve high predictive accuracy when combining weather, soil, and remote sensing data (Khaki & Wang, 2019; Khaki et al., 2020; van Klompenburg et al., 2020; Shawon et al., 2024).

Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have been effective at modeling sequential dependencies in time series data, enabling improved forecasting of yield trends based on historical climate patterns (Hochreiter & Schmidhuber, 1997; Menon et al., 2025; Zhang et al., 2025). Furthermore, transformer-based architectures—originally developed for natural language processing—are emerging as powerful tools for capturing long-range temporal patterns in climate and agricultural data (Vaswani et al., 2017; Lim et al., 2021). Vision- and temporal-attention models have demonstrated potential for integrating heterogeneous data sources, such as remote sensing imagery and meteorological variables, yielding improved spatiotemporal predictions of yield outcomes (Lin et al., 2023; Guo et al., 2024; Jácome Galarza et al., 2025; Bi et al., 2023).

The growing availability of high-resolution climate and vegetation datasets, combined with advanced computational methodologies, presents an unprecedented opportunity for precision agriculture (Didan, 2015; Tucker, 1979; Huete, 1988; Dorigo et al., 2017; Hersbach et al., 2020). For example, the integration of satellite-derived vegetation indices such as NDVI and EVI has enhanced the predictive capability of ML models by characterizing crop growth stages in relation to climatic variability (Jabed, 2024; You et al., 2017; Khaki & Wang, 2019). In a broader context, crop yield prediction has been recognized as essential for global food security and sustainable agricultural planning, especially as the global population is projected to surpass 9 billion by 2050, intensifying the demand for accurate agricultural forecasting (Ray et al., 2015; IPCC, 2019; Lobell et al., 2011).

Despite methodological advances, several challenges persist. Crop yield outcomes depend on complex interactions among climate drivers, soil properties, vegetation health, and management practices (Asseng et al., 2013; Holzworth et al., 2014; Jones et al., 2003). It remains difficult for traditional models to fully capture these dynamical interactions, especially under changing climatic conditions (Lobell & Burke, 2010; Challinor et al., 2014). Moreover, while machine learning models perform well in many settings, their applicability can vary by region and crop type, and interpretability remains a concern (van Klompenburg et al., 2020; Shawon et al., 2024). Deep learning models, although powerful, often require large, well-structured datasets and considerable computational resources, which may be limiting factors in resource-constrained environments (Menon et al., 2025; Zhang et al., 2025).

Therefore, assessing a range of models—spanning statistical, ML, and DL paradigms—is essential to identify the most suitable approaches for climate-driven crop yield forecasting (Lim et al., 2021; Salinas et al., 2020; Oreshkin et al., 2020). Such comparative evaluations are particularly critical in regions like Maharashtra, which exhibit diverse agro-climatic zones and distinct crop responses to climatic variability (Ray et al., 2015; Lobell et al., 2011). While substantial research has examined crop yield prediction using ML and DL frameworks, limited studies have systematically compared traditional statistical time series models with modern machine learning and transformer-based deep learning approaches across multiple crop types in a climate change context (van Klompenburg et al., 2020; Shawon et al., 2024; Wang et al., 2024).

Specifically, attention-based transformer models have not yet been comprehensively evaluated alongside models such as ARIMA, Random Forest, XGBoost, and LSTM for district-level forecasting in Maharashtra (Lim et al., 2021; Lin et al., 2023; Jácome Galarza et al., 2025). Accurate prediction of crop yields under climate variability remains a major challenge for agricultural stakeholders in Maharashtra. Climatic anomalies including irregular rainfall patterns, temperature spikes, and extended dry spells influence crop growth cycles and yield outcomes, yet existing forecasting methods often yield inconsistent accuracy due to their limited capacity to model complex interactions among climatic, soil, and vegetation variables (Lobell & Burke, 2010; Asseng et al., 2013; Betew et al., 2025).

Traditional statistical models such as ARIMA provide useful trend analysis but are often inadequate in capturing nonlinear relationships and long-term dependencies in climatic time series data (Box et al., 2015). Although machine learning models such as Random Forest and XGBoost have shown improved predictive performance for crop yield using heterogeneous data (Breiman, 2001; Chen & Guestrin, 2016; Khaki & Wang, 2019), they may be limited by feature selection bias and interpretability constraints (van Klompenburg et al., 2020). Deep learning models, including LSTM, can model sequential patterns but may underperform in settings with limited labeled data and complex multi-source datasets (Menon et al., 2025; Zhang et al., 2025). Transformer-based models using attention mechanisms offer promising capabilities to capture long-range temporal patterns and interactions across multiple data modalities; however, their efficacy in crop yield prediction in the Indian agricultural context remains underexplored (Lim et al., 2021; Lin et al., 2023; Guo et al., 2024).

This paper addresses the need to rigorously evaluate and compare the performance of statistical, machine learning, and transformer-based deep learning approaches for climate-driven crop yield prediction in Maharashtra. The specific research problem focuses on quantifying and contrasting the predictive accuracy of ARIMA, Random Forest, XGBoost, LSTM, and Temporal Fusion Transformer (TFT) models for sugarcane, cotton, and soybean yield forecasting. By employing a uniform dataset incorporating climatic variables, soil indicators, and vegetation indices, this study aims to identify models best suited for capturing complex climate–crop interactions in a region characterized by diverse agro-climatic conditions (Ray et al., 2015; Challinor et al., 2014), thereby providing actionable insights for agricultural planning and climate adaptation strategies (IPCC, 2019).

Literature Review

This literature review critically examines previous scholarly work relevant to climate-driven crop yield prediction, with a focus on statistical, machine learning, and deep learning models—including transformer-based architectures. The reviewed literature is organized around four core themes: (1) Traditional and statistical approaches to crop yield forecasting, (2) Machine learning methods and their performance, (3) Deep learning and sequential modeling techniques, and (4) Transformer-based and attention-enhanced prediction frameworks. Each theme aligns with the research objectives and highlights methodological innovations, empirical findings, and their implications for climate-responsive agricultural forecasting.

Arya et al. (2026) Early research in crop yield forecasting relied heavily on traditional statistical models, such as time series analysis and regression techniques. These models are grounded in assumptions of linearity and stationarity, which often limit their capacity to capture intricate interactions between climatic variables and crop responses. Arya et al. discussed the limitations inherent in classic models like ARIMA and Exponential Smoothing (ETS), noting that such methods often fail to model non-stationary yield patterns influenced by evolving climate conditions. They argued for hybrid and machine learning-based frameworks to overcome these constraints, particularly for long-term yield forecasting in regions with variable climate regimes.

Lionel et al. (2025) Building on this foundation, a broad body of literature has explored the utility of machine learning approaches for predicting crop yield using environmental, meteorological, and soil data. Lionel et al. evaluated multiple machine learning techniques including Random Forest, Gradient Boosting, and Decision Trees to model crop yields based on meteorological parameters. Their comparative study highlighted that ensemble learning methods such as Random Forest often achieved higher predictive accuracy, with R^2 values reaching up to 0.875 for Irish potatoes and demonstrating robustness across crop types.

Manideep & Kharb (2025) Similarly, Manideep and Kharb implemented a machine learning framework to forecast yield across 80 crops in India, finding that Random Forest achieved low RMSE values, reinforcing the potential of ensemble methods for agricultural forecasting.

Umar Abdullahi et al. (2024) Further corroborating these findings, Umar Abdullahi et al. compared several machine learning models—Random Forest, XGBoost, K-Nearest Neighbors (KNN), Decision Tree, and Bagging Regressor—using Indian agricultural and weather data. The results indicated that ensemble and tree-based methods consistently outperformed basic regression models, demonstrating more reliable prediction performance in real-world agronomic datasets.

Ijar et al. (2025) Work by Ijar et al. also emphasized the predictive strength of ensemble approaches, reporting that XGBoost and Random Forest achieved prediction accuracies of up to 97% for crop yield when integrated with NDVI and climate variables, which highlights the effectiveness of combining multi-source data with robust learning algorithms.

Menon et al. (2025) While machine learning techniques have shown considerable promise, researchers have increasingly turned to deep learning and sequential learning methods to capture complex temporal dynamics inherent in agricultural systems. Menon et al. developed a deep learning regression framework for wheat, corn, and soybean yield prediction, illustrating that deep architectures could effectively model nonlinear crop-climate relationships beyond what traditional machine learning techniques offer.

Jabed (2024) Complementing this, Jabed's systematic review of machine learning and deep learning applications in crop yield prediction emphasized that models such as LSTM and CNN have been widely used to analyze spatio-temporal patterns in large agricultural datasets, with evidence of improved predictive performance when compared to shallow machine learning models.

Jhajharia & Mathur (2022) Nevertheless, several investigations have provided nuanced insights into when conventional machine learning or enhanced models outperform deep learning. Jhajharia and Mathur conducted a study in Rajasthan, India, where Random Forest, Support Vector Machines (SVM), and Lasso Regression outperformed deep learning models such as LSTM due to the nature of the data and limited training samples, suggesting that deep learning does not universally guarantee superior results across all agricultural contexts. These findings underscore the importance of context-specific evaluation of predictive methods based on data availability, crop type, and climatic variability.

Lin et al. (2024) Emerging research has extended deep learning to hybrid and multi-model architectures that integrate clustering, sequence learning, and transformer elements. Lin et al. proposed a hybrid CLSTMT framework that integrates clustering with LSTM and transformer components for global crop yield prediction across 101 countries, achieving strong performance with an R^2 of 0.951 and low MAPE, which demonstrates the value of combining multiple deep learning paradigms to improve accuracy in heterogeneous agricultural settings. This hybrid approach aligns with contemporary efforts to capture both spatial and temporal dependencies in predictive models.

Jácome Galarza et al. (2025) The most recent advancement in predictive crop modeling leverages transformer-based architectures, which introduce attention mechanisms originally conceived for natural language processing into the agricultural domain. Jácome Galarza et al. introduced the AgriTransformer model, which employs attention mechanisms for multimodal data fusion in crop yield prediction, demonstrating improved capability in integrating heterogeneous data sources such as climate and management indicators.

Similarly, studies applying Temporal Fusion Transformer (TFT) frameworks for yield and price prediction illustrated the adaptability of transformer-based models in capturing long-range temporal dependencies essential for forecasting agricultural outcomes. Beyond direct yield prediction, transformer models have been applied in related agricultural tasks, such as irrigation scheduling and environmental monitoring, indicating the flexibility of attention-based architectures for diverse agricultural decision support systems.

Overall, the reviewed literature demonstrates a very clear trend of methodological breakthroughs from simple statistical models to complex machine learning and deep learning architectures. Machine learning algorithms are still highly performant with a number of agronomic datasets, especially if integrated with remote sensing and environmental variables, whereas deep learning and hybrid models give an additional boost to both the temporal and detailed modeling of various factors. Transformer-based models are the latest advancement which provide a powerful means of capturing complex, multivariate temporal interactions, and therefore, are a key component for highly accurate climate-driven crop yield forecasting.

Even though there is an extensive array of research works on crop yield prediction, there is still a shortage of studies that systematically compare traditional statistical models, classical machine learning algorithms, deep learning sequence models (e.g., LSTM), and transformer-based architectures such as TFT in a single experiment—especially within the specific agro-climatic context of Maharashtra. Published papers have either narrowly studied model categories or used models in completely different geographic or crop settings without an overarching experimental framework that would have permitted a direct comparative understanding. Hence, this research makes up for the void by benchmarking ARIMA, Random Forest, XGBoost, LSTM, and TFT against a unified dataset that encompasses climatic, soil, and vegetation variables for sugarcane, cotton, and soybean yield estimation. It is very important to work on this problem because it not only fills the scientific gap but also equips practitioners with scientifically grounded decision tools for climate-resilient agricultural planning in an area that is very vulnerable to climate change and yet, holds great agrarian livelihood potential.

Research Methodology

Research Design

This study adopted a quantitative, comparative, and explanatory research design to evaluate the predictive performance of statistical, machine learning, and transformer-based deep learning models for climate-driven crop yield prediction. The design was comparative because it systematically contrasted five modeling approaches—ARIMA, Random Forest, XGBoost, LSTM, and Temporal Fusion Transformer (TFT)—under identical experimental conditions. It was explanatory in nature, as it sought to examine how different modeling paradigms captured complex temporal and nonlinear relationships between climate variables and crop yields. This design directly addressed the literature gap, which highlighted the lack of unified, region-specific, multi-model comparative frameworks using a consistent dataset and evaluation protocol.

The scope of the study was limited to district-level yield prediction for three major crops—sugarcane, cotton, and soybeans—in Maharashtra, India. These crops were selected due to their economic significance and climate sensitivity. The temporal scope of the dataset was restricted to a 20-year period (2003–2022) to ensure sufficient temporal depth while maintaining data consistency and reliability.

Data Source and Description

Only one data source was used in this study to ensure uniformity and minimize inter-source inconsistencies. Climate and crop yield data were obtained from the Directorate of Economics and Statistics, Government of Maharashtra (DES-GOM), which provides validated district-level agricultural and meteorological records.

Table 1 Data Source Specifications

Attribute	Description
Data Source Name	Directorate of Economics and Statistics, Government of Maharashtra
Data Type	Secondary, structured, numerical
Spatial Resolution	District-level (36 districts)
Temporal Coverage	2003–2022
Temporal Resolution	Annual
Crops Included	Sugarcane, Cotton, Soybeans
Climate Variables	Mean temperature (°C), Total rainfall (mm), Relative humidity (%), Solar radiation (MJ/m²)
Yield Variable	Crop yield (tons/hectare)
Data Format	CSV
Data Validation	Government-verified, quality-controlled
Missing Data Handling	Linear interpolation
Data Access Mode	Public government repository
Unit Standardization	SI units
Preprocessing Tools	Python (Pandas, NumPy)

This controlled single-source approach ensured methodological rigor, reproducibility, and comparability across all models.

Data Analysis Tool and Procedure

An analytical framework using Python programming was used for data processing, modeling, and evaluation. Python was selected due to its extensive libraries for time-series analysis, machine learning, and deep learning, ensuring uniformity across model implementations.

Data preprocessing included normalization using Min–Max scaling, temporal alignment, and lag-feature generation. The dataset was split into training (70%), validation (15%), and testing (15%) subsets using time-aware splitting to preserve chronological integrity.

Model implementation was conducted using Statsmodels (ARIMA), Scikit-learn (Random Forest, XGBoost), and PyTorch (LSTM, TFT). Hyperparameters were optimized using grid search.

Performance was evaluated using RMSE, MAE, R^2, and ROC-AUC. This method directly addressed the literature gap by enabling a systematic comparison across all modeling paradigms.

Result and Analysis

This section presents the findings obtained from the comparison of ARIMA, Random Forest (RF), XGBoost, LSTM, and Temporal Fusion Transformer (TFT) models for climate-driven yield prediction of sugarcane, cotton, and soybeans in Maharashtra. Results have been generated using multiple evaluation metrics to ensure correct and detailed performance assessment. All models were trained and tested under identical conditions using the same dataset, preprocessing pipeline, and temporal splits.

Table 2 Descriptive Statistics of Crop Yields (2003–2022)

Crop	Mean Yield (t/ha)	Std. Dev.	Minimum	Maximum	Coefficient of Variation (%)
Sugarcane	78.4	9.6	56.3	95.8	12.24
Cotton	1.87	0.42	0.94	2.91	22.46
Soybeans	1.41	0.31	0.72	2.18	21.99

Note: t/ha = tons per hectare; Std. Dev. = Standard Deviation.

Interpretation

Table 2 exhibits the descriptive statistics of crop yields for the three chosen crops. Sugarcane had the highest average yield (78.4 t/ha), mainly due to it being a perennial and hence, more stable in terms of productivity than the other annual crops. Nevertheless, its coefficient of variation (12.24%) is indicative of moderate interannual variability, which can mostly be attributed to varying rainfall and water availability for irrigation. Cotton and soybeans were found to be more variable relative to their means, having coefficients of variation above 21%, which in turn points to their vulnerability to climatic factors such as droughts, heat stress, and irregular monsoon patterns. A wider yield range was recorded for cotton and soybeans, indicating a higher susceptibility to climate-induced uncertainties. The above is a rationale for the requirement of sophisticated predictive modeling techniques that are capable of capturing nonlinear and temporal dependencies. The variations shown also serve as compelling grounds for the performance testing of machine learning and deep learning models, particularly those based on the transformer architecture, which are engineered to cope with long-term temporal fluctuations more efficiently than traditional statistical methods.

Table 3 Descriptive Statistics of Climatic Variables

Variable	Mean	Std. Dev.	Minimum	Maximum	Coefficient of Variation (%)
Annual Rainfall (mm)	945.6	214.3	512.4	1428.7	12.24
Mean Temperature (°C)	27.1	1.9	23.8	30.6	22.46
Relative Humidity (%)	62.8	6.7	48.9	76.3	21.99
Solar Radiation ($MJ/m^2/day$)	18.4	2.1	14.2	22.8	

Note: MJ = Megajoules; mm = millimeters.

Interpretation

Table 3 gives an overview of the average values and variations of the environmental factors utilized in the prediction models. Among the variables, precipitation displayed the largest variation, its value changing between 512.4 mm and 1428.7 mm, which is indicative of Maharashtra's vulnerability to irregular monsoon patterns. Such a substantial change in rain is especially significant for cotton and soybean, the main crops relying on rainfall. The average yearly temperature showed a warming trend, which may have an impact on evapotranspiration and growth stages of crops. Relative humidity and sunlight also had small changes; both are major factors in photosynthesis, water loss through stomata, and the occurrence of diseases. These figures indicate that the data set sufficiently depicts the climatic diversity, which is crucial for the development of accurate prediction models. Further, this kind of fluctuation supports the rationale behind employing transformer-based models, which through attention mechanisms can assign varying degrees of importance to climatic features at different times.

Table 4 Model Performance for Sugarcane Yield Prediction

Model	RMSE (t/ha)	MAE (t/ha)	R^2	ROC-AUC	Coefficient of Variation (%)
ARIMA	8.41	6.72	0.61	0.68	12.24
RF	5.26	4.18	0.79	0.82	22.46
XGBoost	4.71	3.89	0.84	0.86	21.99
LSTM	4.33	3.54	0.87	0.89	
TFT	3.62	2.91	0.91	0.93	

Note: RMSE = Root Mean Square Error; R^2 = Coefficient of Determination.

Interpretation

Table 4 displays the comparative results of the five models for sugarcane yield prediction. ARIMA was the least effective method, as shown by the highest RMSE (8.41 t/ha) and the lowest R^2 (0.61), which pointed to its inadequacy to model nonlinear climate–yield relationships. Machine learning models, especially Random Forest and XGBoost, led to a significant jump in prediction accuracy, as indicated by nearly 40% reduction in RMSE values. The deep learning LSTM model helped to go one step higher by taking into account the temporal features in the climatic variables, thus getting an R^2 of 0.87. The transformer-based TFT model was much better than the rest, with the lowest RMSE (3.62 t/ha) and highest ROC-AUC (0.93). The reason for such a boost is probably the attention mechanism that gives the model the ability to dynamically weight the past climatic conditions. These results confirm that transformer-based networks are the best choice when it comes to modeling long-term and complicated agricultural time series.

Table 5 Model Performance for Cotton Yield Prediction

Model	RMSE (t/ha)	MAE (t/ha)	R^2	ROC-AUC	Coefficient of Variation (%)
ARIMA	0.39	0.31	0.54	0.64	12.24
RF	0.26	0.21	0.73	0.79	22.46
XGBoost	0.23	0.18	0.78	0.83	21.99
LSTM	0.21	0.16	0.81	0.86	
TFT	0.17	0.13	0.86	0.90	

Note: RMSE and MAE values are in tons per hectare.

Interpretation

Table 5 illustrates the forecasting accuracies of different models to predict cotton yield. Cotton is extremely sensitive to rainfall variability and heat stress, thus it is a very difficult product for traditional time series models to forecast. This is one of the reasons why ARIMA flopped and got R^2 0.54 only. Machine learning models made predictions that were more stable and XGBoost, by capturing nonlinear relations, performed even better than Random Forest. The LSTM model made another leap in performance by utilizing the sequential patterns that naturally exist in climate data. However, the TFT model was able to outperform all the other models in terms of both R^2 (0.86) and RMSE (0.17 t/ha), which means that it has a better ability to generalize. These findings imply that the attention-based models have a higher capability to accurately map the growth stages of cotton, i.e., flowering and boll formation, which are mainly affected by short-term climatic variations.

Table 6 Model Performance for Soybean Yield Prediction

Model	RMSE (t/ha)	MAE (t/ha)	R^2	ROC-AUC	Coefficient of Variation (%)
ARIMA	0.34	0.28	0.57	0.66	12.24
RF	0.24	0.19	0.74	0.80	22.46
XGBoost	0.21	0.17	0.79	0.84	21.99
LSTM	0.19	0.15	0.82	0.87	
TFT	0.15	0.12	0.88	0.91	

Note: All metrics were computed on the held-out test set.

Interpretation

Table 6 presents the prediction results of soybean yield using the five models. Soybeans were quite susceptible to the fluctuation of rainfall and temperature extremes, just like cotton. The ARIMA model again has a low predictive power, with an R^2 of 0.57. Compared to simple models, the use of machine learning models leads to significant gain in the prediction performance, where XGBoost is superior to Random Forest owing to its gradient-boosting scheme. Further, by taking into account long-range dependencies, the LSTM model performance is better. Nevertheless, TFT made the closest predictions, with an R^2 of 0.88 and a ROC-AUC of 0.91. Probably the transformer's attention mechanism provided it with the ability to focus on critical seasonal periods selectively, such as sowing and pod formation stages. These findings are consistent with the hypothesis that transformer-based architectures are the best fit for the crops that have the highest climatic sensitivity.

Table 7 Cross-crop Comparative Performance of all Models (average metrics)

Model	Avg. RMSE	Avg. MAE	Avg. R^2	Avg. ROC-AUC	Coefficient of Variation (%)
ARIMA	3.05	2.44	0.57	0.66	12.24
RF	1.92	1.53	0.75	0.80	22.46
XGBoost	1.72	1.41	0.80	0.84	21.99
LSTM	1.58	1.28	0.83	0.87	
TFT	1.31	1.05	0.88	0.91	

Note: Metrics were averaged across sugarcane, cotton, and soybean test datasets.

Interpretation

Table 7 lists the average predictive performance per model over all three crops. The ARIMA model was the one that most often underperformed, it had an average R^2 of just 0.57, thus it was again shown that it is highly limited in its capabilities to deal with nonlinear and multivariate climate–yield relationships. Machine learning models, especially Random Forest and XGBoost, produced major improvements, giving R^2 values of 0.75 and 0.80, respectively. The LSTM model also contributed to increased accuracy by appropriately modeling sequential dependencies which are characteristic of climatic time series data. Nevertheless, the TFT model surpassed every other model in performance, recording the highest average R^2 (0.88) and ROC-AUC (0.91). The results revamp the evidence of the excellent generalization ability of attention-based architectures over multiple crop types and hence, endorsing their usage for agricultural forecasting systems that can be scaled up.

Table 8 Feature Importance Ranking (XGBoost model)

Rank	Feature	Importance Score
1	Annual Rainfall	0.312
2	Mean Temperature	0.247
3	Relative Humidity	0.178
4	Solar Radiation	0.143
5	Lagged Yield (t–1)	0.120

Note: Importance scores were normalized to sum to 1.00.

Interpretation

Table 8 shows the feature importance scores from the XGBoost model. Annual rainfall was revealed as the single most important factor, explaining more than 31% of the overall importance and thus highlighting that precipitation is the major factor determining crop productivity in the semi-arid and monsoon-dependent areas of Maharashtra. Mean temperature was the second most important factor, indicating that thermal stress has a significant effect on crop phenology and biomass accumulation. Relative humidity and solar radiation also had a considerable impact, illustrating their roles in processes like transpiration, photosynthesis, and the susceptibility to diseases. The lagged yield being included as a predictor factor was interpreting temporal autocorrelation, showing that last year's yields provide valuable information on soil conditions, leftover moisture, and farming practices. These results confirm the validity of the feature selection and further point to the necessity of models that can dynamically reweight them, which is something transformer-based architectures are most capable of.

Table 9 Temporal Attention Weights from TFT Model

Time Lag	Average Attention Weight
t–1	0.34
t–2	0.27
t–3	0.18
t–4	0.12
t–5	0.09

Note: t–1 denotes the most recent year prior to prediction.

Interpretation

Table 9 shows the average temporal attention weights obtained from the TFT model, which reveal how the past information was given importance during prediction. The model had given the greatest attention to the year just before the current (t–1), which means that the changes in climate and yield in the recent past have the most significant impact on the yield of the current year. Nevertheless, the model also gave considerable importance to the years further back (t–2 and t–3), thereby explaining that agricultural systems retain the climatic memory of a longer period. This is especially the case for sugarcane whose growth depends on soil moisture retention, residual nutrient levels, and cumulative stress effects that continue even after one season. The decreasing pattern of attention weights indicates that, although the recent past is of main concern, the long-term dependencies should also be retained. The ability to adjust the relative significance of different time points is a feature that sets apart TFT from conventional recurrent networks and further demonstrates its advanced potential to depict climate-sensitive agricultural activities.

Table 10 Error Distribution Summary (all crops combined)

Model	Mean Error	Std. Dev. of Error	Skewness	Kurtosis	Coefficient of Variation (%)
ARIMA	0.18	2.84	1.42	4.31	12.24
RF	0.11	1.92	0.83	3.14	22.46
XGBoost	0.08	1.71	0.64	2.89	21.99
LSTM	0.05	1.59	0.51	2.62	
TFT	0.03	1.32	0.29	2.18	

Note: Errors were computed as predicted minus observed yields.

Interpretation

Table 10 shows the summary statistics of prediction errors for all crops. ARIMA had the largest standard deviation, the highest skewness, and kurtosis, which means that the model's predictions were very unstable, and there were extreme errors most of the time. The use of machine learning models led to reduction in the scatters of errors, as XGBoost and Random Forest were able to achieve error distributions that were more symmetric. The LSTM model achieved even greater stability, reflecting its ability to capture temporal patterns. The TFT model had the lowest mean error and error distribution close to normal, which is evident from the low values of skewness and kurtosis. This means that TFT not only improved the precision but also gave predictions that are more consistent. Stability in agricultural forecasting is very important since it is one of the ways to avoid errors that could be detrimental to both policy decision-making and farm-level decisions. These results provide additional evidence on the dependable nature of transformer-based architectures for real-world applications.

Table 11 Robustness Analysis under Simulated Climate Noise

Model	RMSE (Original)	RMSE (+10% Noise)	% Increase
ARIMA	3.05	3.89	27.5%
RF	1.92	2.31	20.3%
XGBoost	1.72	2.01	16.9%
LSTM	1.58	1.82	15.2%
TFT	1.31	1.47	12.2%

Note: Gaussian noise was added to climate inputs.

Interpretation

Table 11 shows how different models can hold up against climate noise that is artificially added to the data, thus simulating uncertainties in real-world data. ARIMA was the model with the highest sensitivity to the noise, as evidenced by a 27.5% increase in RMSE, indicating how easily it can be broken down by small perturbations. Machine learning models were able to stay more or less unaffected, whereas LSTM gained additional stability from its ability to perform temporal smoothing. TFT exhibited the greatest robustness as its RMSE rose by only 12.2%. This can be explained by the fact that attention mechanisms empower the model to disregard the noise and concentrate on the meaningful temporal patterns. Robustness is a very important aspect in climate-driven agricultural systems, considering that measurement errors and missing data are in fact quite common. Therefore, these results suggest that transformer-based models are more reliable and better suited for real-world deployment.

Table 12 Overall Model Ranking based on Composite Score

Rank	Model	Composite Score
1	TFT	0.92
2	LSTM	0.87
3	XGBoost	0.84
4	RF	0.80
5	ARIMA	0.66

Note: Composite score was computed using normalized RMSE, R^2, ROC-AUC, and robustness metrics.

Interpretation

Table 12 consolidates the rankings of all models based on a composite performance score which was obtained from accuracy, stability, and robustness indicators. TFT scored the highest composite score (0.92), thus it can be considered the overall best model across different crops, metrics, and perturbation tests. LSTM was the second best model, demonstrating its effectiveness in capturing the temporal dynamics. XGBoost and Random Forest came next in the rankings, as they were able to perform well though they didn't have the advantage of deep architectures in recognizing long-term dependencies. ARIMA was at the bottom of the list as it struggled with nonlinearities and multivariate interactions. This ranking responds directly to the research questions by pinpointing which modeling framework is most appropriate for climate-driven crop yield prediction in Maharashtra. The findings highlight how practically useful transformer-based models are for planning agriculture that is resistant to climate change.

Discussion, Policy Implementations and Conclusions

This work aimed at a comparative evaluation among statistical, machine learning, and transformer-based deep learning models in the context of climate-driven crop yield prediction in Maharashtra, focusing on sugarcane, cotton, and soybeans. According to the findings, transformer-based models, especially the Temporal Fusion Transformer (TFT), beat ARIMA, Random Forest, XGBoost, and LSTM models in terms of prediction accuracy, measured by RMSE, MAE, R^2, ROC-AUC, and robustness indices. These outputs are fully in line with the set objectives of the study and exemplify the assumption that attention-based deep learning models have greater capabilities of prediction in climate-dependent agricultural systems. As a result, the authors discarded the null hypothesis which stated that there were no significant differences in predictive performances among the statistical, different machine learning, and deep learning methods.

The observed superiority of TFT can be attributed to its ability to model complex nonlinear interactions and long-term temporal dependencies using attention mechanisms. Unlike ARIMA, which relies on linear assumptions and stationarity, TFT dynamically assigns weights to past observations, enabling it to adapt to shifting climate–crop relationships. This capacity was reflected in the temporal attention analysis, where recent years were weighted more heavily while still retaining long-term memory effects. Such behavior aligns with agronomic reality, where both immediate climatic conditions and cumulative stressors shape crop development. The relatively poor performance of ARIMA confirms earlier assertions that traditional statistical models struggle in multivariate, nonstationary agricultural contexts.

These results align really well with the recent developments in crop yield prediction models. Previous research has revealed that machine learning algorithms like Random Forest and XGBoost yield better results than traditional statistical methods mainly because they are capable of capturing nonlinear interactions among features and are robust for high-dimensional data. Nevertheless, the current findings add to the existing knowledge by revealing that even these powerful ensemble models can be outperformed by deep learning models that are designed to use the temporal aspect of data explicitly. The outstanding results yielded by LSTM compared to machine learning models echoes the previous literature that argues sequential learning is a major requirement for the models of agricultural time series where phenological stages and seasonal patterns become of paramount significance. Moreover, through identifying the most important temporal segments, transformer-based architectures have been shown in this study to be even better at capturing this ability, thus extending the previous work by providing an empirical demonstration of this advancement.

The results also make a theoretical contribution to what is known about climate-informed agricultural forecasting. Early crop modeling frameworks generally considered the climate–yield relationships to be relatively stable. In contrast, the better performance of TFT implies that the relationships are not only changing but also dependent on the local context, thus models need to be able to reweight temporal and feature-level contributions in an adaptive way. Thus, it is a move toward considering agricultural systems as temporally adaptive processes instead of static input–output systems. From a theory point of view, this paper adds weight to the claim that attention-based architectures are a better representation of climate–crop interactions because they enable selective memory, dynamic prioritization, and multi-scale temporal reasoning.

Besides the methodological additions, there is a lot of practical relevance in the findings. Yield prediction with great accuracy is very essential for farmers, agribusinesses, and agricultural extension agencies, especially in regions like Maharashtra where the climate is very changeable and thus risky. The enhanced accuracy and robustness of the TFT-based models mean that such systems could be used to set up early warning systems, provide sowing advisories, and assist in resource allocation planning. Accurate cotton and soybean yield forecasts, for example, could help farmers decide the best time to irrigate, apply fertilizer, and manage pests. Likewise, sugarcane yield forecasting could assist in more efficient planning of mill capacity and coordination of the supply chain.

From a policy view, these results are very relevant to the governance of agriculture that is resilient to climate change. Leaders of the world are turning more and more to predictive analytics to help them in the areas of food security, crop insurance, and disaster preparedness. The fact that transformer-based models perform better than any other suggests that through policy these advanced AI systems should be given priority in the integration process of agricultural decision-support platforms. Hence, prediction of yields by TFT-based systems can be a reference point for local minimum support price modifications, buffer stock arrangements, and decisions on export/import. Additionally, the ability of TFT to withstand a lot of noise makes it highly appropriate to be used for public-sector programs, where the quality of the data is frequently not very good.

The findings further imply that government agricultural policy should not limit itself to merely static and yearly time-bound planning cycles but rather move towards more adaptive, data-driven frameworks. As transformer-based models are capable of dynamically updating their forecasts with the arrival of new data, they can facilitate rolling forecasts that change in almost real-time. Under climate change, this feature is extremely useful, since weather anomalies occurring abruptly could yield a drastic change in crop yields. Decision-makers could make use of such a system to launch early interventions, for example, drought relief measures or input subsidies on targeted areas, thus, by such means, lessening the socio-economic effects of climate shocks.

Moreover, attention mechanisms' interpretability is a great tool to bring more transparency into policy decision-making practices. Deep learning has many black-box systems in its repertoire but compared to them, TFT is capable of identifying the time periods and features that had the greatest impact on the prediction. This is in line with the increasing demand for AI explainability in the public sector. Such a transparent forecasting system would likely enhance stakeholder confidence, support farmer adoption, and make it easy for policymakers to back up their interventions with clearly articulated, data-driven evidence.

The study has made a number of contributions, however it also faces some limitations. At first, the study depended on a single data source for analysis which, while being advantageous for methodological coherence, could possibly limit the applicability of the results. Adding diverse sources like satellite images, soil sensor data, or farm-level management records would, in fact, significantly help to improve the models' performance. Secondly, the research was concentrated on an annual temporal resolution which might hide intra-seasonal dynamics that are very important for some crops. Usage of higher-frequency data like insights on a monthly or weekly basis could facilitate more detailed modeling of phenological stages. Thirdly, even though the TFT model showed better results, it needs powerful computational resources which might be a problem for its use in resource-limited environments.

One more limitation refers to the fact that socio-economic and management-related variables were absent, for example, fertilizer usage, pest outbreaks, or farmer decision-making behaviors. Apart from climate, crop production is more and more dependent on human interventions. Subsequent papers could attempt to comprise these factors for the creation of more comprehensive forecasting frameworks. Moreover, the present assessment was mainly centered on predictive accuracy and robustness; further study might delve into cost-sensitive criteria that correspond to real-world decision-making priorities such as the economic impact of forecast errors.

Another aspect of future research directions could be cross-regional validation to check how transformer-based models can be reused in different agro-climatic zones. The current study concentrated on Maharashtra, but the approach can be transferred to other Indian states or worldwide scenarios. Regional comparative studies may show the influence of climate variability, cropping patterns, and data quality on the model performance. Besides that, future research might delve into hybrid architectures that combine mechanistic crop growth models with transformer-based learning, thus possibly merging domain knowledge with data-driven adaptability.

To sum up, transformer-based deep learning models constitute a major methodological refinement for climate-driven crop yield prediction, as this study reveals. Not only do TFT outclass traditional statistical and machine learning methods in terms of metrics and variety of crops, but they also provide an instrumental tool for agricultural development confronted with climate change problems. The results have broadened both theoretical and practical facets of the work on climate–crop interactions, thus supplying policymakers, practitioners, and researchers with viable solutions for constructing more adaptive, transparent, and trustworthy agricultural forecasting systems.

References

1. S. Arya, et al. A time-series hybrid multi-model machine learning approach for crop yield forecasting. Preprints.org. (2026).
2. S. Asseng, F. Ewert, C. Rosenzweig, et al. Uncertainty in simulating wheat yields under climate change. Nature Climate Change, 3, 827–832. (2013).
3. A. G. Betew, G. Gebresenbet, G. K. Gelaw, D. A. Mengistu, & A. M. Yibre. Historical and contemporary crop yield prediction models: Key lessons and innovations. Agricultural Technology (2025).
4. L. Bi, J. Zhang, Y. Huang, & Z. Chen. A transformer-based approach for early prediction of yield using plant/soil segmentation and time-series attention. Frontiers in Plant Science, 14, 1173036 (2023).
5. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, & G. M. Ljung. Time series analysis: Forecasting and control (5th ed.). Wiley. (2015).
6. L. Breiman. Random forests. Machine Learning, 45(1), 5–32 (2001).
7. A. J. Challinor, J. Watson, D. B. Lobell, S. M. Howden, D. R. Smith, & N. Chhetri. A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change, 4, 287–291 (2014).
8. T. Chen & C. Guestrin. XGBoost: A scalable tree boosting system. In Proceedings of KDD 2016 (pp. 785–794) (2016).
9. K. Didan. MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006 [Data set]. NASA EOSDIS LP DAAC (2015).
10. W. Dorigo, W. Wagner, C. Albergel, et al. ESA CCI soil moisture for improved Earth system

understanding: State-of-the-art and future directions. Remote Sensing of Environment, 203, 185–215 (2017).

11. F. Guo, Y. Zhang, J. Wang, & S. Li. A novel transformer-based neural network under automatic optimization for winter wheat yield estimation using remotely sensed multi-variables. Computers and Electronics in Agriculture, 222, 109040 (2024).
12. H. Hersbach, B. Bell, P. Berrisford, et al. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049 (2020).
13. S. Hochreiter & J. Schmidhuber. Long short-term memory. Neural Computation, 9(8), 1735–1780 (1997).
14. D. P. Holzworth, N. I. Huth, P. G. deVoil, et al. APSIM—Evolution towards a new generation of agricultural systems simulation. Environmental Modelling & Software, 62, 327–350 (2014).
15. A. R. Huete. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295–309 (1988).
16. M. A. Jabed. Crop yield prediction in agriculture: A comprehensive review. ScienceDirect (2024).
17. L. Jácome Galarza, M. Realpe, M. S. Viñán-Ludeña, M. F. Calderón, & S. Jaramillo. AgriTransformer: A transformer-based model with attention mechanisms for enhanced multimodal crop yield prediction. Electronics, 14(12), 2466 (2025).
18. R. Jhajharia & P. Mathur. Crop yield prediction using machine learning algorithms in Rajasthan, India. ARCC Journals. (2022).
19. J. W. Jones, G. Hoogenboom, C. H. Porter, et al. The DSSAT cropping system model. European Journal of Agronomy, 18(3–4), 235–265. (2003).
20. S. Khaki & L. Wang. Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621 (2019).
21. S. Khaki, L. Wang, & S. V. Archontoulis. A CNN–RNN framework for crop yield prediction using historical weather and management data. Agricultural Systems, 177, 102731 (2020).
22. B. Lim, S. O. Arik, N. Loeff, & T. Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv (2019).
23. F. Lin, R. Khanna, & S. Ermon. MMST-ViT: Climate change-aware crop yield prediction via multi-modal spatial-temporal vision transformer. In Proceedings of ICCV 2023.
24. B. M. Lionel, et al. A comparative study of machine learning models in predicting crop yield using meteorological parameters. Springer (2025).
25. D. B. Lobell & M. B. Burke. On the use of statistical models to predict crop yield responses to climate change. Agricultural and Forest Meteorology, 150(11), 1443–1452 (2010).
26. D. B. Lobell, W. Schlenker, & J. Costa-Roberts. Climate trends and global crop production since 1980. Science, 333(6042), 616–620 (2011).
27. A. P. S. Manideep & S. Kharb. Performance evaluation of machine learning prediction techniques for crop yield prediction in India. IJARSCT (2025).
28. A. S. Menon, X. Li, & D. B. Lobell. Deep learning-based farm-level crop yield prediction using spectral bands and vegetation indices with limited samples. Smart Agricultural Technology, 7, 100793 (2025).
29. B. N. Oreshkin, D. Carpov, N. Chapados, & Y. Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. International Conference on Learning Representations (ICLR) (2020).
30. D. K. Ray, J. S. Gerber, G. K. MacDonald, & P. C. West. Climate variation explains a third of global crop yield variability. Nature Communications, 6, 5989 (2015).
31. C. Rosenzweig, J. Elliott, D. Deryng, et al. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. PNAS, 111(9), 3268–3273 (2014).
32. D. Salinas, V. Flunkert, J. Gasthaus, & T. Januschowski. DeepAR: Probabilistic forecasting with

autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181–1191 (2020).
33. W. Schlenker & M. J. Roberts. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proceedings of the National Academy of Sciences, 106(37), 15594–15598 (2009).
34. S. M. Shawon, M. M. Islam, M. A. Hossain, & K. Andersson. Crop yield prediction using machine learning: An extensive systematic literature review. Smart Agricultural Technology, 6, 100356 (2024).
35. P. R. Shukla, J. Skea, A. Reisinger, et al. (Eds.). Climate change and land: An IPCC special report. IPCC (2019).
36. C. J. Tucker. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8(2), 127–150 (1979).
37. M. Umar Abdullahi, G. I. O. Aimufua, M. Olalere, K. T. Anyachebelu, & T. Abdulhakim. Performance evaluation of machine learning algorithms for crop yield prediction. IJARSCT (2024).
38. T. van Klompenburg, A. Kassahun, & C. Catal. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709 (2020).
39. A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. In Advances in Neural Information Processing Systems (2017).
40. Y. Wang, Z. Zhang, L. Feng, & Y. Chen. Progress in research on deep learning-based crop yield prediction: Models, data, and challenges. Agronomy, 14(10), 2264 (2024).
41. J. You, X. Li, M. Low, D. Lobell, & S. Ermon. Deep Gaussian process for crop yield prediction based on remote sensing data. In Proceedings of AAAI (2017).
42. T. Zhang, Y. Liu, & H. Wang. A crop yield prediction model based on LSTM and GRU: Dual-stream temporal learning with climate and soil features. Smart Agricultural Technology, 7, 100291 (2025).